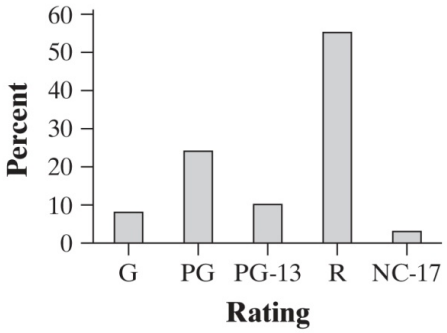
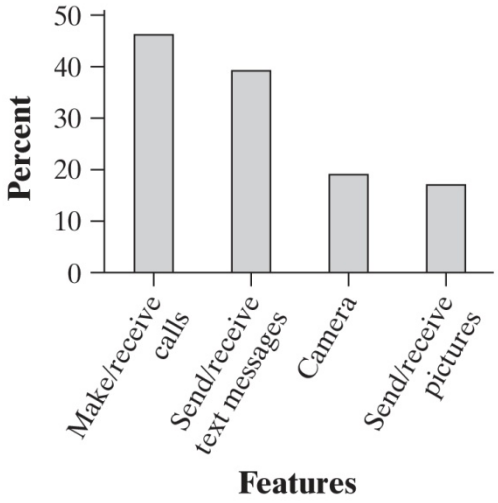
Chapter Review Exercises (page 76)

R1.1 (a) The individuals are movies.
(b) The variables are Year (quantitative), Rating (categorical), Time (quantitative), Genre (categorical), and Box office sales (quantitative). *Note*: Year might be considered categorical if we want to know how many of these movies were made each year rather than the average year.
(c) This movie is Avatar, released in 2009. It was rated PG-13, runs 162 minutes, is an action film, and had box office sales of $2,781,505,847.

R1.2 A bar chart is given below.



R1.3 (a) The graph is misleading because the "bars" are different widths. For example, the bar for "send/receive text messages" should be roughly twice the size of the bar for "camera" when it is actually much more than twice as large.
(b) It would not be appropriate to make a pie chart for these data because they do not describe parts of a whole. Students were free to answer in more than one category.
(c) A bar graph is given below.



R1.4 (a) Of the 219 who responded to the survey, $78 + 49 + 21 = 148$ were Facebook users. So, 148/219 = 67.6% were Facebook users. This is part of a marginal distribution because it is part of the

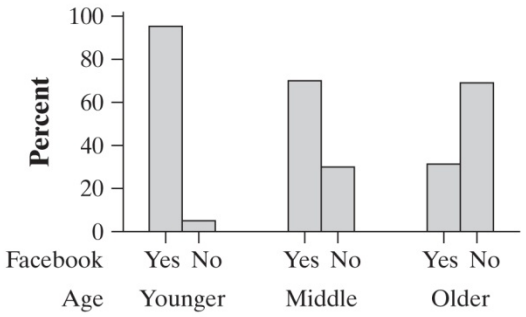
distribution of one variable for all categories of the other variable. In other words, it uses only column and table totals.

(b) There were $78 + 4 = 82$ younger students and 78 of those were Facebook users. So 78/82 = 95.1% of the younger students were Facebook users. There were 148 Facebook users, so 78/148 = 52.7% of the Facebook users were younger.

R1.5 There does appear to be an association between age and Facebook status. Looking at the conditional distributions of Facebook status for each of the age groups, we get the following table:

	Facebook user?	
Age	Yes	No
Younger (18-22)	95.1%	4.9%
Middle (23-27)	70.0%	30.0%
Older (28 and up)	31.3%	68.7%

This can also be seen graphically in the bar graph given below:



From both the table and the graph we can see that as age increases the percent of Facebook users decreases. For younger students, about 95% are members. That drops to 70% for middle students and drops even further to 31.3% for older students.

R1.6 (a) A stemplot is shown below.

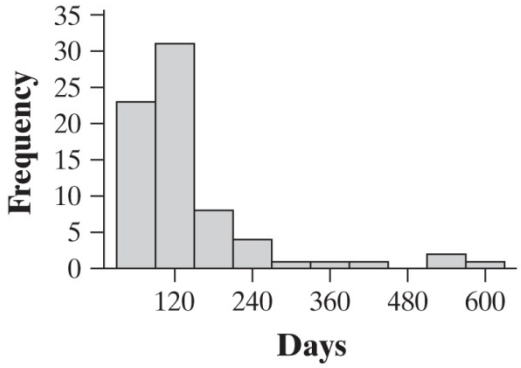
```
48 | 8
49 |
50 | 7
51 | 0
52 | 6799
53 | 04469
54 | 2467
55 | 03578
56 | 12358
57 | 59
58 | 5
```

Key: 48 | 8 = 4.88

(b) The distribution is roughly symmetric with one possible outlier at 4.88. The center of the distribution is between 5.4 and 5.5. The densities vary from 4.88 to 5.85.

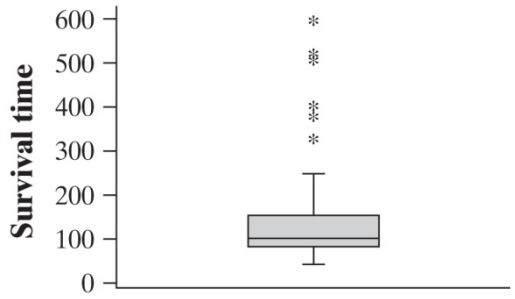
(c) Since the distribution is roughly symmetric, we can use the mean to estimate the Earth's density to be about 5.45 times the density of water.

R1.7 (a) A histogram is shown below.



The survival times are right skewed, as expected. The median survival time is 102.5 days and the range of survival times is $598 - 43 = 555$ days. There are several high outliers with survival times above 500.

(b) The boxplot is shown below:



(c) Because the distribution is strongly skewed, it would be better to use the median and *IQR* to summarize the distribution. The outliers will have a big effect on the mean and standard deviation.

R1.8 (a) About 20% of low-income and 33% of high-income households consisted of two people.

(b) The shapes of both distributions are skewed to the right, however the skewness is much stronger in the distribution for low income households. On average, household size is larger for high income households. In fact, the majority of low-income households consist of only one person. Only about 7% of high-income households consist of one person. One-person households might have less income because they would include many young single people who have no job or retired single people with a fixed income.

R1.9 (a) The amount of mercury per can of tuna will typically vary from the mean by about 0.3 ppm.

(b) The $IQR = 0.380 - 0.071 = 0.309$, so any point below $0.071 - 1.5(0.309) = -0.393$ or above $0.38 + 1.5(0.309) = 0.8435$ would be considered an outlier. Because the smallest value is 0.012, there are no low outliers. According to the histogram, there are values above 0.8435, so there are several high outliers.

(c) The distribution of the amount of mercury in cans of tuna is highly skewed to the right. The median is 0.18 ppm and the *IQR* is 0.309 ppm.

R1.10 The albacore tuna generally has more mercury. Its minimum, first quartile, median and third quartile are all greater than the respective values for light tuna. But that doesn't mean that light tuna is always better. It has a much bigger spread of values with some cans having as much as twice the amount of mercury as the largest amount in the albacore tuna. The distribution for light tuna is skewed to the right with several high outliers. The distribution for albacore tuna is more symmetric with just a couple of high outliers.